ESIP Earth Science Data Analytics (ESDA) Cluster

October 23, 2014

Agenda

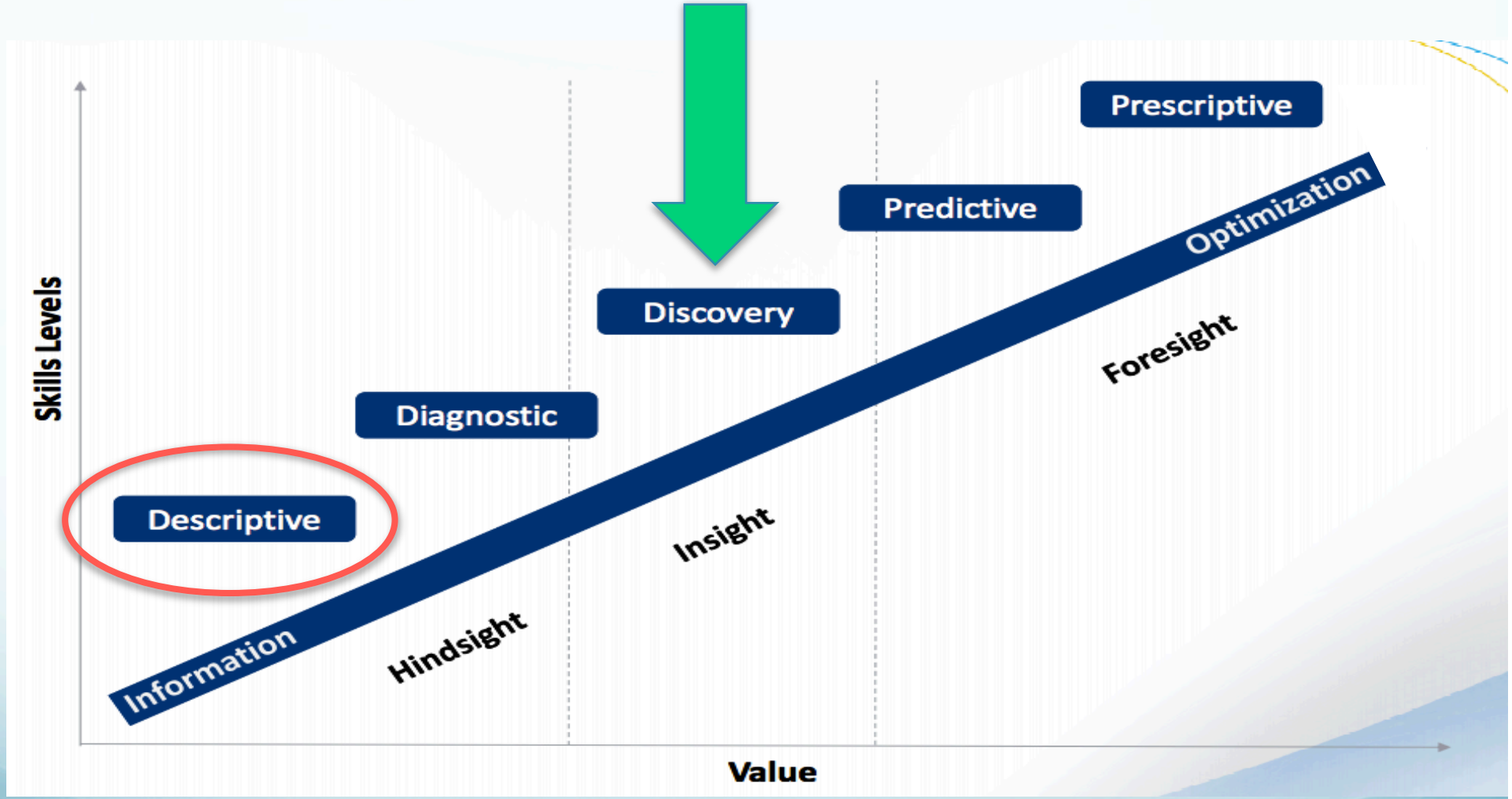
- 5 min – **Recap of our last telecon on Descriptive Analytics**
- 15 min - **Guest Speakers: George Djorgovski**, Cal Tech, who is interested in the roles of computation in knowledge discovery.
- 20 minutes – **Discussion: Diagnostic Analytics** (and how it compares to Descriptive Analytics)
- 15 minutes – **Planning ahead discussion**:
  - December Breakouts: Suggestions for guest speakers
  - Are we starting to learn enough to write a paper on the Types of Data Analytics Utilized in (the various data analysis phases of) Earth Science

Frisco Recap – Actions

- Flush out use cases: Bound the issue, be specific. Seek additional individuals who are facing issues utilizing large heterogeneous datasets
- Further define Data Analytics types: Per type: Issues, Potential solutions, exemplary situations, user classes, other
- Initiate some of the above planned mapping
- In January, have 2 ESDA sessions. One can be entitled: 'Earth Science Data Analytics 101'

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion - Descriptive Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Defining - Descriptive Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

What does Descriptive Data Analytics mean? What does it do? How it is used? Examples! Where in Earth science would this be used? Which users?

- purpose of descriptive analytics is to summarize and tell you what has happened in the past
- "the simplest class of analytics," one that allows you to condense big data into smaller, more useful nuggets of information. *http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-2-Understanding-Predictive-Analytics/ba-p/79616*
- compute descriptive statistics (i.e. counts, sums, averages, percentages, simple arithmetic) that summarizes certain groupings or filtered version of the data, which are typically simple counts of some events. They are mostly based on standard aggregate functions *http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-1-Descriptive-Analytics/ba-p/77766*

Now For Our Guest Speaker:
Dr. George Djorgovski

Shamelessly taken from Dr. Djorgovski's on-line CV:

Dr. George Djorgovski

Professor of Astronomy, California Institute of Technology

Relevant Professional Interests: Development of e-Science/Cyber-infrastructure, the roles of computation in knowledge discovery, Astroinformatics, Virtual Observatory, advanced data-mining and exploration techniques.

Defining - Descriptive Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

What does Descriptive Data Analytics mean? What does it do? How it is used? Examples! Where in Earth science would this be used? Which users?

- The purpose of descriptive analytics is simply to summarize and tell you what happened. For example, number of post, mentions, fans, followers, page views, kudos, +1s, check-ins, pins, etc. ...simple event counters.
- Other descriptive analytics may be results of simple arithmetic operations, such as share of voice, average response time, % index, average number of replies per post, etc.
  http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-1-Descriptive-Analytics/ba-p/77766
- Descriptive analytics is simple, all we need is data
- following the NetFlix approach, Amazon uses "Descriptive" analytics to process what you have purchased in the past, to predict what books, videos, and things you might like in the future

Defining - Descriptive Analytics

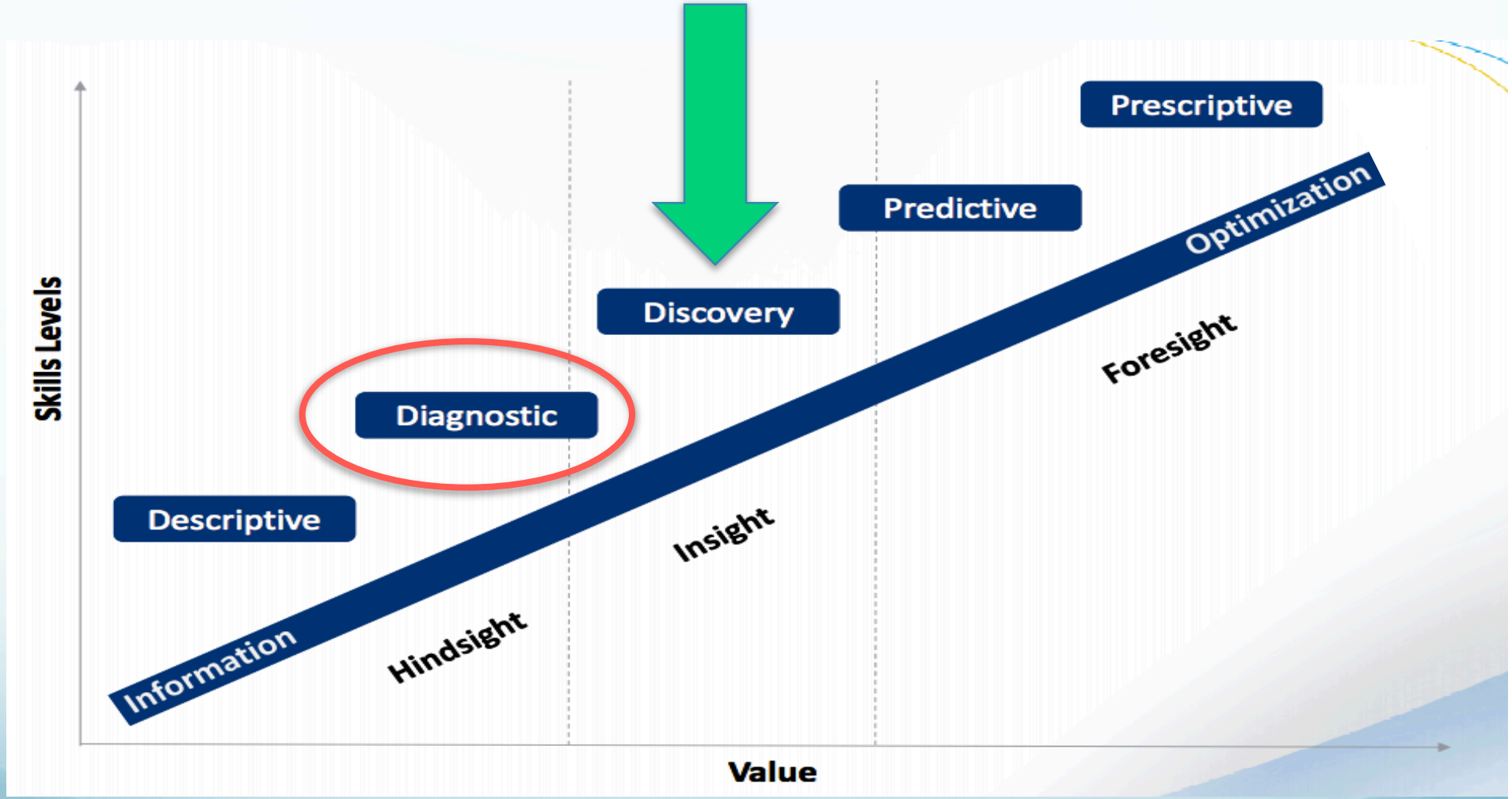
Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

What does Descriptive Data Analytics mean? What does it do? How it is used? Examples! Where in Earth science would this be used? Which users?

- Descriptive analytics answers the question, "What happened...?" It looks at data and information to describe the current situation in a way that trends, patterns and exceptions become apparent http://www.mu-sigma.com/analytics/ecosystem/dipp.html
- Descriptive statistics is the discipline of quantitatively describing the main features of a collection of information or the quantitative description http://en.wikipedia.org/wiki/Descriptive_statistics
- Natural Hazards: Looking for Patterns and Trends; Bringing in heterogeneous datasets, together summarized, to detect patterns
- Erin to provide slides: air quality 'use case'

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion – Diagnostic Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Discussion – Diagnostic Analytics

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Diagnostic analytics
Determine *why* something happened, using content analytics and natural language processing to cull insights found in documents, email, websites, social media and so on. Understand the root cause of geophysical changes through more detailed analysis and visualizations.
(modified from: http://www.ibm.com/analytics/us/en/analytics-tools.html)

Diagnostic analytics looks deeper into what has happened and seeks to understand why a problem or event of interest occurs. How do various measurable events and actions in the focal domain relate to each other? (http://www.lifescaleanalytics.com/files/lifescale/files/brief_descriptivetoprescriptive.pdf)

Diagnostic data analytics is used to answer the question "Why is it happening?". It strives to identify root causes, key factors, and unseen patterns (
http://webcache.googleusercontent.com/search?q=cache:abyglyZBFLIJ:www.ag-ai.nl/download/17445-21-3-art.Parekh.pdf+&cd=8&hl=en&ct=clnk&gl=us&client=safari)

Descriptive vs. Diagnostic Analytics

Descriptive Analytics	Diagnostic Analytics
You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.	If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use ... intelligence tools to get some insights
better understand what has happened	explain why an event happened
What has happened? How many? How much? Is this changing over time? The objective is to quantify, track and report what might have previously been only a vague qualitative sense for how things are going	looks deeper into what has happened and seeks to understand why a problem or event of interest occurs. How do various measurable events and actions in the focal domain relate to each other? While it may start with bivariate relationships it progress into development of multi-variable explanatory models
performed on historical data to establish statistical benchmarks in order to gain insights or answer the question ”What is happening?”	is used to answer the question “Why is it happening?”

Can we, as a cluster, divide and conquer the Data Analytics Puzzle ...

1. Take what we learn, refine, and define about the different types of Data Analytics

    - Descriptive Analytics
    - Diagnostic Analytics
    - Discoveritive Analytics
    - Predictive Analytics
    - Prescriptive Analytics

2. Associate exemplary Earth science use cases to each type

3. Associate Data Analytics techniques/tools to each type

4. Associate user categories to each type

5. Describe skills and expertise needed for each type

- Currently, we talk about our expertise and experience, but they seldom seem to connect to each other

- This will help us, the industry, and hopefully, educators, focus their understanding and interests regarding Earth Science Data Analytics.

User Model (Subsetted from ESDSWG WG)

Classes	Definition
Public	interested user of no or limited scientific skill
Graduate student	person of moderate to high skill at a university or college working towards an advanced degree
Production Centers	large organization that handles/processes vast quantities of data
Science Team	group of scientists focused on a specific area of study or on a specific instrument type, can include cal/val scientists
QA/Testing	developers or scientists using data to test software operation or to determine quality of a product, can include cal/val scientists
Data Analyst	person using NASA data to perform a specific analysis.
Domain Scientist	person using data to do research and publish within a discipline, comes in with some expertise in using the data
Interdisciplinary Scientist	person using high-level data products from multiple sources
Operational User	Data analyst or tech using data for operational support (applications) and emergency response
Assimilation Modelers	persons or groups that routinely obtain vast quantities of data for incorporation into models, can have operational needs

Relevant AGU Sessions

- Teaching Science Data Analytics Skills Needed to Facilitate Heterogeneous Data/Information Research: The Future Is Here - Session ID#: 1879
- Identifying and Better Understanding Data Science Activities, Experiences, Challenges, and Gaps Areas - Session ID#: 1809
- Advancing Analytics using Big Data Climate Information System - Session ID#: 3022
- Big Data in the Geosciences: New Analytics Methods and Parallel Algorithm - Session ID#: 3292
- Leveraging Enabling Technologies and Architectures to enable Data Intensive Science - Session ID#: 3041
- Open source solutions for analyzing big earth observation data - Session ID#: 3080
- Technology Trends for Big Science Data Management - Session ID#: 2525